



DOCUMENT RESUME

ED 236 202

TM 830 711

AUTHOR Fuchs, Douglas; And Others
TITLE Systematic Bias in the Assessment of Handicapped Children.
INSTITUTION Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY Office of Special Education and Rehabilitative Services (ED), Washington, DC.
REPORT NO IRLD-RR-134
PUB DATE Jul 83
CONTRACT 300-80-0622
NOTE 38p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Education; *Examiners; Experimenter Characteristics; Interaction; *Language Handicaps; Preschool Education; Rapport; *Test Anxiety; *Test Bias; *Testing; Test Validity
IDENTIFIERS *Familiarity; Testing Conditions

ABSTRACT

Prior research demonstrates that examiner unfamiliarity negatively affects the optimal performance of handicapped preschoolers. The present investigation sought to determine whether examiner unfamiliarity also interferes with the optimal performance of handicapped school-age pupils and nonhandicapped children. Sixty-four subjects (16 language-handicapped and 16 nonhandicapped preschoolers and 16 language-handicapped and 16 nonhandicapped school-age students) were tested twice during a period of 2 weeks, once by a familiar examiner and once by an unfamiliar examiner, within a crossover design. A significant interaction was obtained for examiner familiarity and handicapped status, indicating that whereas nonhandicapped subjects scored similarly when tested by familiar and unfamiliar examiners, handicapped children scored higher with the familiar tester. Thus, findings indicated that examiner unfamiliarity negatively affects both language-handicapped preschool and school-age children's performance relative to a normative population; it appears to depress selectively handicapped children's test performance, thereby indicating that an examiner's unfamiliarity constitutes a negatively, systematically biasing condition and threatens the validity of handicapped students' test performance. (Author)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

SYSTEMATIC BIAS IN THE ASSESSMENT OF
HANDICAPPED CHILDREN

Douglas Fuchs, Lynn S. Fuchs, Maryann H. Power, and Ann M. Dailey



SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned
this document for processing
to:

In our judgement, this document
is also of interest to the clearing-
houses noted to the right. Index-
ing should reflect their special
points of view.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

J. Yseeldyke

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

**Institute for
Research on
Learning
Disabilities**

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.



Director: James E. Ysseldyke

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with Special Education Programs, Department of Education. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of Special Education Programs.

Research Report No. 134

SYSTEMATIC BIAS IN THE ASSESSMENT OF
HANDICAPPED CHILDREN

Douglas Fuchs

Department of Education
Clark University

Lynn S. Fuchs

Wheelock College

Maryann H. Power and Ann M. Dailey

Department of Communication Disorders
Worcester State College

July, 1983

Abstract

Prior research demonstrates that examiner unfamiliarity negatively affects the optimal performance of handicapped preschoolers. The present investigation sought to determine whether examiner unfamiliarity also interferes with the optimal performance of handicapped school-age pupils and nonhandicapped children. Sixty-four subjects (16 language-handicapped and 16 nonhandicapped preschoolers and 16 language-handicapped and 16 nonhandicapped school-age students) were tested twice during a period of 2 weeks, once by a familiar examiner and once by an unfamiliar examiner, within a crossover design. A significant interaction was obtained for examiner familiarity and handicapped status, indicating that whereas nonhandicapped subjects scored similarly when tested by familiar and unfamiliar examiners, handicapped children scored higher with the familiar tester. Thus, findings indicated that examiner unfamiliarity negatively affects both language-handicapped preschool and school-age children's performance relative to a normative population; it appears to depress selectively handicapped children's test performance, thereby indicating that an examiner's unfamiliarity constitutes a negatively, systematically biasing condition and threatens the validity of handicapped students' test performance.

Systematic Bias in the Assessment of Handicapped Children

There is general and chronic disregard for the importance of situational variables in the assessment of children. This is true whether the purpose of testing is for educational programming (Labov, 1973; Mehan, 1978; Sigel, 1975) or research (Bronfenbrenner, 1979; Cazden, 1973; Cole & Bruner, 1972; Mishler, 1979). Nevertheless, a limited but growing body of empirical evidence demonstrates that systematic variation of certain typical features of the assessment situation predictably affects test performance (Sattler, 1974).

The effect of the situational factor "examiner familiarity" has been explored relatively frequently. Interest in this factor often has been based on the long-standing developmental notion that children derive much of their comprehension and feelings about a situation from significant adults in that situation (cf. Freud, 1921-22; Piaget, 1965). A recent literature review (Fuchs & Fuchs, 1983) identified 20 investigations of examiner familiarity effects. Twelve of these studies reported subjects performing significantly better under the familiar condition. Of this group, 10 investigations involved handicapped, minority, and/or preschool subjects, thereby suggesting the relative importance of examiner familiarity to children with one or more of these characteristics. Handicapped preschoolers' test performance appeared to increase most consistently in the familiar examiner condition. Among six pertinent studies using handicapped preschoolers as subjects, five (92%) investigations reported positive, significant results.

By demonstrating that handicapped preschoolers, and other select groups, perform differently with familiar and unfamiliar examiners,

recent research has identified a source of error in the manner in which tests commonly are administered. Findings indicate that the typical testing procedure discourages handicapped preschoolers' optimal, absolute performance. A related and more important issue is whether examiner unfamiliarity negatively influences these children's performance relative to some normative populations; i.e., whether the strangeness of a tester selectively depresses the performance of handicapped preschool children. If so, a tester's unfamiliarity would be biased systematically against, and would threaten the validity of, handicapped preschool children's performance. The present study explored this possibility by employing both handicapped and nonhandicapped preschool children.

This study also sought to extend previous research along a second, heretofore unexplored dimension: namely, whether school-age handicapped children, as well as handicapped preschoolers, demonstrate differential performance in favor of the familiar tester. By examining preschool and school-age children with a similar handicapping condition, this study explored whether handicapped preschoolers' sub-optimal performance with strange testers is, for the most part, a developmental issue, or whether it is more closely related to the fact that they are handicapped. Nonhandicapped school-age children also were incorporated into the design to address the possibility of a systematically biasing effect of examiner unfamiliarity at the school-age level, as well as among preschoolers. Finally, in an effort to shed light on the possibly mediating roles of CA and handicapping conditions in examiner familiarity/unfamiliarity

effects, testers' ratings of children's test-related behaviors were collected and analyzed.

Method

Subjects

A total of 64 subjects comprised four different groups of equal size: 16 handicapped preschoolers (HP), 16 handicapped school-age children (HS), 16 nonhandicapped preschoolers (NP), and 16 nonhandicapped school-age children (NS). The HPs and HSs were moderately to profoundly speech- and/or language-impaired, performed within the normal range on individually administered IQ tests, and were participating in one of two language programs (preschool or elementary school levels) that were administered by the same public educational collaborative. The admission criteria of these special education programs included performance on speech and/or language measures that was at least $1\frac{1}{2}$ standard deviations below the mean. NPs and NSs were drawn from a large college-affiliated nursery school and public elementary school, respectively. All subjects were Caucasian, English-speaking, and from predominantly middle-class families located in five contiguous towns in Central Massachusetts. The mean CA for handicapped and nonhandicapped subjects was 77.44 (SD = 24.91) and 76.91 (SD = 24.49) months, respectively; average CA for preschool and school-age subjects was 57.16 (SD = 7.20) and 97.19 (SD = 18.48) months, respectively. A two-way analysis of variance (ANOVA; Handicapped vs. Nonhandicapped and Preschool vs. School-age) revealed a significant difference between the CAs of preschool and school-age subjects, $F(1,60) = 128.28$, $p < .001$; there was no significant

disparity in CA between handicapped and nonhandicapped children, $F(1,60) = .02$, ns, and there was no significant interaction, $F(1,60) = .00$, ns. Additionally, identical numbers of male ($N = 20$) and female ($N = 12$) subjects constituted the handicapped, nonhandicapped, preschool, and school-age groups.

Examiners

There were 32 examiners. All were Caucasian, female, and trained as speech clinicians at a state college in Central Massachusetts. Sixteen examiners were seniors who were one month away from graduation and who had completed a year-long practicum in clinical or educational settings. The remaining 16 testers were enrolled in a Masters program in Communication Disorders and were practicing speech clinicians. Table 1 is a summary of the examiners in terms of whether they tested handicapped, nonhandicapped, preschool, or school-age subjects, and reports means and standard deviations for the four groups' CA and professional experience at the preschool and school-age levels. Two-way ANOVAs (Handicapped vs. Nonhandicapped and Preschool vs. School-age) revealed no significant differences between the examiner groups on any of these variables. (See Table 1 for F and p values.) Additionally, a multivariate two-way analysis of variance indicated there were no differences between the examiner groups in terms of their attitudes toward and conceptualization of handicapped vs. nonhandicapped adults or children as measured on the Role Category Questionnaire (Crockett, 1965; see Fuchs, Fuchs, Dailey, & Power, 1983, for details on administration and scoring).

Insert Table 1 about here

Design

Each examiner was matched randomly with one of the four study sites, resulting in the combination of four seniors and four professionally experienced speech clinicians per subject group. Each of eight examiners per site was assigned randomly to four subjects, two with whom they became personally familiar and two to whom they remained strangers. In this way, examiners served in both familiar and unfamiliar roles, thereby controlling for potentially confounding effects of testers' personality.

With every examiner testing four children, each of the 16 subjects was assessed twice during a period of two weeks, once by a familiar examiner and once by the unfamiliar examiner within a crossover design. Examiners assessed one familiar and one unfamiliar subject on one occasion and another familiar/unfamiliar pair of subjects on a second day. On their first day of testing, one-half of the examiners first tested familiar children, then unfamiliar children; the remaining examiners tested their examinees in reverse order. On the second day, examiners who previously had tested initially familiar subjects were required to test unfamiliar children first; those examiners who, on their first day of testing, had assessed unfamiliar children first were required on their second day to test them after familiar examinees. The time of day that subjects were assessed also was controlled; if a child was tested by the

familiar examiner at 9:00 a.m., this child also was examined at 9:00 a.m. by the unfamiliar tester. Finally, testing at each site was conducted in a quiet setting that was familiar to the children.

Procedure

Selection of nonhandicapped subjects. NPs and NSs were selected by a stratified random sampling procedure. Specifically, for every handicapped subject, all nonhandicapped children in the nursery and elementary schools were identified (a) whose CAs were either three months below or above the CA of the handicapped subject, (b) who were of the same gender, and (c) who were not receiving special education services. This generated 32 overlapping groups (16 at preschool level and 16 at school-age level) in which membership ranged from 4 to 9. From each of these groups, two nonhandicapped children were drawn randomly, one designated arbitrarily to participate in the study and the other assigned to a replacement group.

Selection of examiners. Examiners were chosen in a four-stage process. First, an initial pool of 95 prospective examiners was established by identifying those graduate students ($N = 52$) and seniors ($N = 43$) who matriculated in the Communication Disorders Department. Second, the records of these students were examined to determine whether the prospective examiners met three criteria; namely, examiners were required to (a) be female; (b) have a minimum of $\frac{1}{2}$ year experience with preschool children either as a professional speech clinician or as a student-in-training; and (c) have completed successfully a year-long practicum (for seniors) or an advanced diagnostics course (for graduate students). Application of these

criteria reduced the initial pool to 72 students, 35 graduate students and 37 seniors. Third, 22 individuals were chosen randomly from each of these groups, with 16 in each group arbitrarily designated to participate as examiners and 6 in each group assigned to a replacement group. Finally, the 32 prospective examiners were sent letters informing them (a) of an opportunity to participate in an investigation whose "aim is to study the validity of using standardized tests to assess the language performance of various groups of children," (b) of the nature and extent of their responsibility as study participants, and (c) that, in return for their participation, they would be awarded stipends and an opportunity to earn academic credit.

Personal familiarity. Examiners' personal familiarity was induced experimentally by two procedures. First, examiners who were assigned to preschool subjects were required to make one home visit to each of their two "familiar" children. Testers were told that the purpose of these visits was "to get to know the child and to permit the child to get to know you." Examiners also were instructed to bring toys and games with which to play with the children. The home visit was to last about one hour. If subjects were school-age, examiners spoke with their future examinees by telephone instead of visiting them in their homes. Examiners were asked to (a) introduce themselves, (b) discuss briefly the upcoming testing, and (c) inform the children that, prior to testing, there would be an opportunity to play. Testers and children explored together materials that they could bring to this interaction, as well as special activities they

might pursue. Subjects were made to expect the telephone calls via their classroom teachers, and the phone conversations were to last about five minutes.

The study's second strategy to induce personal familiarity required testers to play with their familiar children on a one-to-one basis in a quiet space outside of the subject's classroom. Each tester of preschool children provided the same toys and games with which she and the child had played during the home visit; examiners of school-age subjects attempted to bring materials and/or conduct activities that had been mentioned during their phone conversations. The play sessions always immediately preceded testing and lasted 20 and 30 minutes for preschool and school-age subjects, respectively. Research assistants monitored and occasionally regulated the duration of these play sessions. The lapse in time separating the home visit or phone conversation and testing ranged from two to five days.

Following the data collection phase of this investigation, examiners were asked to complete a questionnaire that was administered during a debriefing session. Their responses indicated that the average home visit (to preschoolers) lasted 60.63 minutes ($SD = 20.70$) and the average phone conversation (with school-age children) lasted 5.81 minutes ($SD = 4.69$). In combination with time spent during the play sessions immediately preceding testing, preschool and school-age subjects spent averages of 80.63 minutes ($SD = 20.70$) and 34.72 minutes ($SD = 7.50$), respectively, with their familiar examiners prior to testing. Handicapped children's total interaction time with familiar testers averaged 56.44 minutes ($SD = 25.97$); nonhandicapped

subjects averaged 58.91 minutes (SD = 29.92) of pretest contact with familiar examiners. A two-way ANOVA (Handicapped vs. Nonhandicapped and Preschool vs. School-age) indicated that, whereas preschool subjects participated in significantly greater amounts of pretest contact with examiners than did school-age subjects, $F(1,60) = 137.57$, $p < .001$, handicapped and nonhandicapped subjects did not differ in this respect, $F(1,60) = .40$, ns.

Measures

The Clinical Evaluation of Language Functions (CELF; Semel-Mintz & Wiig, 1982) is a comprehensive language test that is comprised of two scales, processing (PS) (i.e., auditory comprehension) and production (PC) (i.e., verbal expression). Semel-Mintz and Wiig (1982) reported internal consistency of .82 and a test-retest reliability coefficient of .96. Pilot administrations of the CELF were conducted with language-impaired preschoolers (CAs: 3-10 to 4-8) and nonhandicapped, above-average achieving intermediate grade children (CAs: 9-4 to 12-1) to determine the possibility of "floor" and "ceiling" effects, respectively. Results indicated the CELF was appropriately difficult for the study's diverse sample. Examiners were trained to administer the CELF in one three-hour training session, which was conducted by a certified speech clinician.

Schaeffer's Teacher Rating Scale (STRS: Schaeffer, n.d.) lists 23 traits and their polar opposites that describe students' personality (e.g., nervous/relaxed), sociability (e.g., shy/outgoing), and work habits (e.g., easily distracted, concentrates). Each bipolar dimension is rated along a 7-point scale where 1 and 7 signify

most negative and positive responses, respectively. Each examiner completed the STRS for her two familiar and two unfamiliar examinees during a debriefing session that was conducted one week following the last administrations of the CELF. Internal consistency alpha for the STRS was .97. (The STRS is provided in the Appendix.)

Results

Means and standard deviations on each scale of the CELF are displayed in Table 2, for each examinee group in both familiar and unfamiliar testing conditions.

Insert Table 2 about here

The PS scale on the CELF is comprised of six subtests; the PC scale, four. In order to compare children's performance on the two scales, weighted scores were calculated: the PS scale score was divided by six and the PC scale score by four. Then, a two-between (Handicapped vs. Nonhandicapped and Preschool vs. School age), two-within (Familiar vs. Unfamiliar tester and the CELF scales, PS vs. PC) analysis of variance (ANOVA) was run on the CELF. This ANOVA revealed significant F values for handicapped status, $F(1,60) = 16.98$, $p < .001$, age status, $F(1,60) = 81.18$, $p < .001$, for the handicapped by age interaction, $F(1,60) = 11.98$, $p < .001$, for the CELF scale by age interaction, $F(1,60) = 7.90$, $p < .01$, and for the familiarity condition by handicapped status interaction, $F(1,60) = 4.00$, $p < .05$. Weighted mean scores and standard deviations on the two CELF scales for each of the four examinee groups in both unfamiliar and familiar testing conditions are shown in Table 3.

Insert Table 3 about here

Inspection of Tables 2 and 3 reveals that nonhandicapped students performed better on the CELF than did the handicapped children (mean difference = 37.28 unweighted points or 7.24 weighted points), and that school-aged youngsters scored higher than preschoolers (mean difference = 75.28 unweighted points or 15.84 weighted points). The interaction between the familiarity condition and handicapped status is illustrated in Figure 1 for the weighted CELF scores and suggests that whereas nonhandicapped children scored similarly when tested by familiar and unfamiliar examiners (scoring, on the average, 0.05 points lower on each subtest with the familiar tester), handicapped youngsters scored higher when tested by the familiar examiner (scoring an average 1.86 points better with the familiar tester on each subtest).

Insert Figure 1 about here

Additional significant interactions indicated that (a) the difference between preschool and school-age children's performances was greater for nonhandicapped than for handicapped youngsters, and (b) that the difference between performances on the PS and PC scales of the CELF was greater for preschool than for school-age children. However, because these two findings are extraneous to the purpose of this paper, they are not discussed further.

For each Schaeffer Rating Scale protocol, the total score was divided by the number of items on the scale to derive an average score per item. On these average scores per item, a two-between (Handicapped vs. Nonhandicapped and Preschool vs. School-age), one-within (Familiar vs. Unfamiliar) ANOVA was run. The ANOVA yielded one significant F value for age status, $F(1,60) = 4.26$, $p < .05$; one additional F value approached significance for handicapped status, $F(1,60) = 3.36$, $p = .07$. No significant effect was obtained for familiarity status, $F(1,60) = .03$, ns . Means and standard deviations on the Schaeffer, for each of the four examinee groups in both unfamiliar and familiar testing conditions are shown in Table 4. These means indicate that school-age students' ratings on the Schaeffer were an average .49 point higher per item than were the ratings of preschoolers, and that nonhandicapped youngsters were rated an average .44 point higher per item than were the handicapped students. Testers' ratings of familiar and unfamiliar subjects were nearly identical.

 Insert Table 4 about here

Discussion

Findings indicate that, in comparison to nonhandicapped children, handicapped subjects performed significantly stronger when tested by an examiner with whom they were personally familiar. This result extends previous research on examiner familiarity in two important ways. First, whereas prior investigations demonstrated that tester

unfamiliarity discourages handicapped preschoolers' optimal, absolute test performance, results from this study show that the strangeness of a tester negatively influences handicapped children's performance relative to a normative population; it appears to depress selectively handicapped children's test performance. This finding suggests that an examiner's unfamiliarity constitutes a negatively, systematically biasing condition and threatens the validity of handicapped students' test performance.

Results from the present investigation also extend the importance of examiner familiarity from preschool to school-age handicapped children. The absence of an interaction between familiarity and CA or among familiarity, handicapping condition, and CA suggests there may be something about a child's handicap, irrespective of CA, that promotes differential test performance. Such a causative influence may originate within the child (e.g., self-perceived vulnerability); however, it also may be rooted in (a) examiners' attitudes and behaviors directed toward handicapped children (see Fuchs et al., 1983; Fuchs, Zern, & Fuchs, in press) or (b) an interaction between the perceptions and actions of examiner and handicapped examinees.

In any case, the apparently negatively biasing nature of examiner unfamiliarity for handicapped children has many implications for educational practice. One such implication involves the level of confidence we may place in the predictive validity of a test with respect to students' classroom performance. Despite the fact that most testers are unfamiliar and classroom teachers are familiar to pupils, predictive validity presupposes a high degree of similarity

between the conditions under which both testing and classroom learning occurs. While this presupposition may hold true for nonhandicapped children, present findings indicate it is not true for handicapped students; the strangeness of an examiner appears to transform the test situation into a qualitatively different experience from that of a classroom setting. Hence, what is learned about a handicapped child in the typical testing environment may have scant relevance for and may be seriously misleading about classroom instruction.

The salience of a strange tester for handicapped, but not for nonhandicapped children, also questions the adequacy of many tests' standardization populations, when such groups are employed to determine handicapped pupils' test performance. Most psychological, language, and educational achievement tests include few if any handicapped children in their normative populations. Regardless of whether these largely nonhandicapped groups were tested by familiar or unfamiliar examiners during the standardization phase, we have no reason to believe their optimal performance was compromised. In contrast, examiner unfamiliarity has been seen to discourage handicapped children's optimal functioning. Thus, when we employ an unfamiliar examiner to assess handicapped pupils and when we compare these children's performance to a typical standardization population, we are comparing handicapped children's suboptimal functioning to the presumably optimal performance of the normative group.

This seemingly "apples and oranges" situation is an explicit violation of PL 94-142 and Section 504, which stipulate that, for handicapped students, test instruments must be selected and

administered to "accurately reflect the student's aptitude or achievement level or whatever other factors the test purports to measure...." (34 C.F.R., §§ 104.35 [b] [3] and 300.532 [c] [3], cited in McCarthy, 1983).

This investigation also explored examiners' ratings of subjects' test behavior on the STRS. Testers described school-age children's behavior as more positive than preschoolers' behavior, and examiners' higher ratings of nonhandicapped than handicapped examinees' behavior approached significance. These results paralleled findings that school-age and nonhandicapped subjects performed more strongly on the CELF than preschool and handicapped children, respectively. Together these findings seem to corroborate prior research demonstrating a positive relation between children's performance on tests and their adaptive, general behavior in the test setting (e.g., Sigel, 1975). In contrast, however, examiners rated familiar and unfamiliar handicapped children's test behaviors similarly, despite the better performance of handicapped children on the CELF in the familiar condition. A previous study (Fuchs et al., 1983), also employing speech clinicians as testers but using a different rating instrument, generated similar results.

It is unclear why testers' ratings of handicapped examinees on the STRS failed to reflect these subjects' stronger CELF performance in the familiar condition. Whatever the reason may be for this seeming insensitivity, there is an important implication: Although many tests depend on examiners to exercise their judgment to determine whether, and if so when, optimal performance has been established,

results from this study and prior research suggest testers have difficulty making this determination, at least with respect to select populations. Further research might explore the utility of more objective procedures, such as a checklist or rating scale that specifies dimensions of rapport, with which testers may determine more accurately a child's valid test performance and his/her level of comfort during an examination.

References

- Bronfenbrenner, U. The ecology of human development. Cambridge, MA: Harvard University, 1979.
- Cazden, C. B. The neglected situation in child language research and education. In F. Williams (Ed.), Language and poverty. Chicago: Rand McNally, 1973.
- Cole, M., & Bruner, J. S. Preliminaries to a theory of cultural differences. In I. Gordon (Ed.), Seventy-first yearbook of the National Society for the Study of Education, Part II -- Early childhood education. Chicago: University of Chicago, 1972.
- Crockett, W. H. Cognitive complexity and impression formation. In B. A. Maher (Ed.), Progress in experimental personality research (vol. 2). New York: Academic Press, 1965.
- Freud, S. [Group psychology and the analysis of the ego]. (J. Strackey, Ed. and trans.). London: International Psycho-Analytical Press, 1922. (Originally published, 1921).
- Fuchs, D., & Fuchs, L. S. Examiner familiarity: A quantitative synthesis. Working paper, 1983.
- Fuchs, D., Fuchs, L. S., Dailey, A. M., & Power, M. H. Effects of pretest contact with experienced and inexperienced examiners on handicapped children's performance (Research Report No. 110). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1983.
- Fuchs, D., Zern, D. S., & Fuchs, L. S. A microanalysis of participant behavior in familiar and unfamiliar test conditions: Exceptional Children, in press.
- Labov, W. The logic of nonstandard English. In F. Williams (Ed.), Language and poverty. Chicago: Rand McNally, 1973.
- McCarthy, M. M. The application of competency testing mandates to handicapped children. Harvard Educational Review, 1983, 53, 146-164.
- Mehan, H. Structuring school structure. Harvard Educational Review, 1978, 48, 32-64.
- Mishler, E. G. Meaning in context: Is there any other kind? Harvard Educational Review, 1979, 49, 1-18.
- Piaget, J. The moral judgment of the child. New York: The Free Press, 1965.

Sattler, J. M. Assessment of children's performance. Philadelphia: W. B. Saunders, 1974.

Schaeffer, S. Schaeffer teacher rating scale. Unpublished manuscript, n.d.

Semel-Mintz, E., & Wiig, E. H. Clinical evaluation of language functions. Columbus, OH: Charles E. Merrill, 1982.

Sigel, I. E. The search for validity or the evaluator's nightmare. In R. A. Weinberg & S. G. Moore (Eds.), Evaluation of educational programs for young children. Washington, D.C.: The Child Development Associate Consortium, 1975.

Footnotes

Douglas Fuchs and Lynn S. Fuchs are also affiliated with the Institute for Research on Learning Disabilities. The research described in the report was supported in part by a grant from the Office of Community Affairs, Division of Graduate and Continuing Education, Worcester State College.

The authors wish to thank Dick Gilmore, Sally Lingner, and Joan Lucey for their administrative support of this investigation. Gratitude for their cooperation also is expressed to the classroom teachers from whose classrooms our subjects were drawn. Finally, we greatly appreciate the efforts of the many speech clinicians who served as familiar and unfamiliar examiners.

Table 1
CA and Experience of Examiners Who Were Assigned to Handicapped,
Nonhandicapped, Preschool, and School-age Subjects^a

| | Handicapped | Nonhandicapped | Preschool | School-age | Factor | ANOVA | |
|--------------------------|----------------|----------------|----------------|----------------|-------------------------|-------|---------|
| | | | | | | df | p-value |
| | 303.00 (62.76) | 276.72 (39.36) | 291.36 (59.88) | 288.00 (47.04) | H/NH ^b | 1,28 | .19 |
| | | | | | P/S ^c | 1,28 | .00 |
| | | | | | H/NH x P/S ^d | 1,28 | .70 |
| School experience | 16.00 (18.20) | 13.88 (15.13) | 14.67 (16.62) | 15.29 (16.96) | H/NH | 1,28 | .71 |
| | | | | | P/S | 1,28 | .88 |
| | | | | | H/NH x P/S | 1,28 | .20 |
| School-age experience | 11.13 (10.48) | 14.26 (16.95) | 9.67 (9.24) | 16.57 (17.80) | N/NH | 1,28 | .65 |
| | | | | | P/S | 1,28 | .20 |
| | | | | | N/NH x P/S | 1,28 | .86 |

Values are means and standard deviations in parentheses, which are calculated in terms of months.

^aHandicap vs. nonhandicap

^bPreschool vs. school-age

^cInteraction between handicap/nonhandicap and preschool/school-age

Table 2

Unweighted Means and Standard Deviations on the CELF Scales for Each Examinee
Group in Familiar and Unfamiliar Testing Conditions^a

| Examinee Group | CELF Scale | | | |
|---------------------------|----------------|----------------|----------------|----------------|
| | PS | | PC | |
| | Familiar | Unfamiliar | Familiar | Unfamiliar |
| Preschool Handicapped | 96.31 (36.62) | 83.38 (38.86) | 36.31 (33.57) | 35.75 (37.22) |
| School-age Handicapped | 139.94 (39.52) | 137.81 (43.26) | 90.94 (44.52) | 75.25 (34.75) |
| Preschool Nonhandicapped | 101.56 (45.64) | 101.31 (40.15) | 45.13 (28.42) | 40.69 (29.23) |
| School-age Nonhandicapped | 211.63 (32.38) | 216.13 (30.39) | 137.81 (51.40) | 138.25 (52.71) |

^aStandard deviations are indicated in parentheses.

Table 3

Weighted Means and Standard Deviations on the CELF Scales for Each
Examinee Group in Familiar and Unfamiliar Testing Conditions^a

| Examinee Group | CELF Scale | | | |
|---------------------------|--------------|--------------|---------------|---------------|
| | PS | | PC | |
| | Familiar | Unfamiliar | Familiar | Unfamiliar |
| Preschool Handicapped | 16.09 (6.08) | 13.90 (6.47) | 9.10 (8.40) | 8.96 (9.31) |
| School-age Handicapped | 23.33 (6.59) | 22.14 (7.23) | 22.76 (11.13) | 18.83 (8.67) |
| Preschool Nonhandicapped | 16.92 (7.61) | 15.84 (5.13) | 9.74 (6.10) | 10.19 (7.32) |
| School-age Nonhandicapped | 35.28 (5.40) | 36.03 (5.08) | 34.48 (12.85) | 34.58 (13.19) |

^aStandard deviations are indicated in parentheses.

Table 4
Average Score Per Item on the Schaeffer for Each Examinee
Group in Familiar and Unfamiliar Testing Conditions

| Examinee Group | Score | | | |
|---------------------------|----------|------|------------|------|
| | Familiar | | Unfamiliar | |
| | Mean | SD | Mean | SD |
| Preschool Handicapped | 4.41 | 0.94 | 4.27 | 1.01 |
| School-age Handicapped | 4.56 | 0.71 | 5.00 | 0.80 |
| Preschool Nonhandicapped | 4.87 | 0.94 | 4.58 | 1.09 |
| School-age Nonhandicapped | 5.36 | 0.82 | 5.18 | 0.95 |

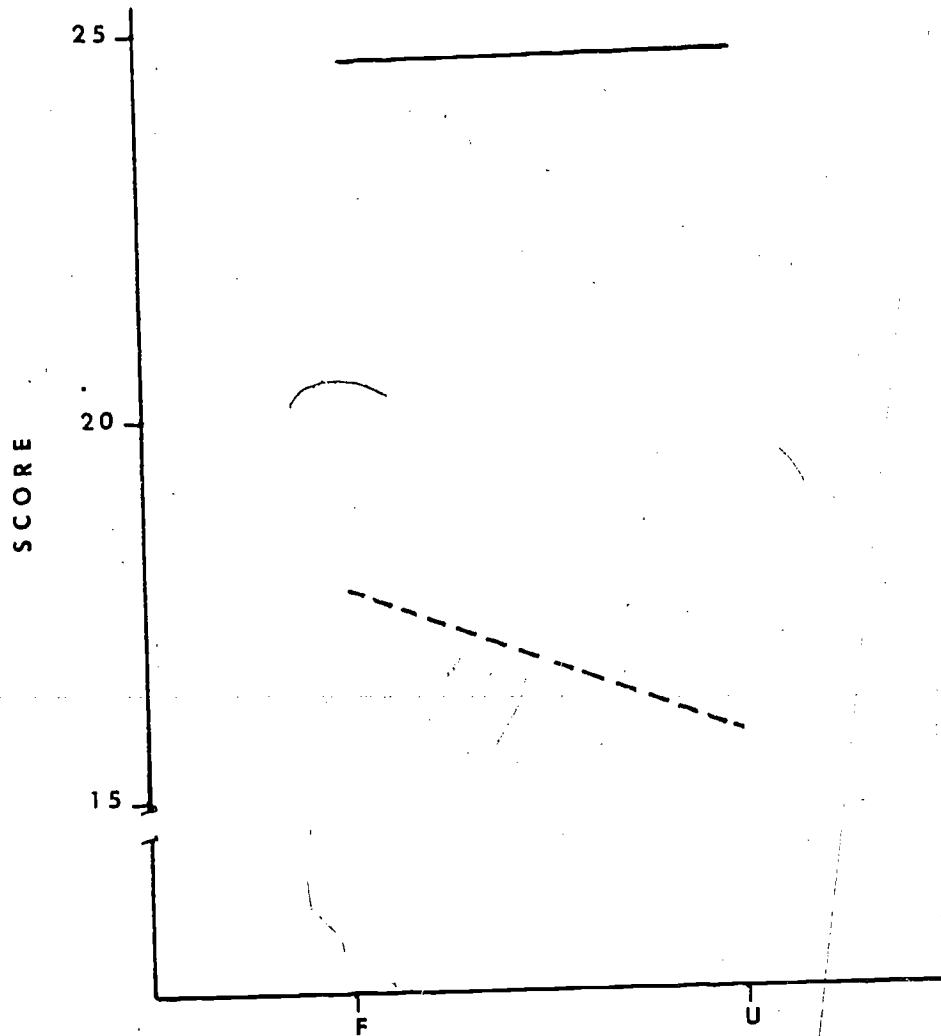


Figure 1. Display of interaction: CELF scores of handicapped (---) and nonhandicapped (—) children in familiar (F) and unfamiliar (U) testing conditions.

APPENDIX

Schaeffer Teacher Rating Scale

Child's Name: _____

Compared with other children the same age, how would you describe this child on the traits listed here. For example, if the item had to do with height, and the child was just a little shorter than average, you might mark it as follows:

Example: short 1 2 ③ 4 5 6 7 tall

Please circle a number on each line below to show where you would rate this child on the following traits:

| | average | | | | | | | |
|---------------------------|---------|---|---|---|---|---|---|-------------------------|
| 1. shy | 1 | 2 | 3 | 4 | 5 | 6 | 7 | outgoing |
| 2. slow learner | 1 | 2 | 3 | 4 | 5 | 6 | 7 | fast learner |
| 3. nervous | 1 | 2 | 3 | 4 | 5 | 6 | 7 | relaxed |
| 4. gloomy | 1 | 2 | 3 | 4 | 5 | 6 | 7 | cheerful |
| 5. short attention span | 1 | 2 | 3 | 4 | 5 | 6 | 7 | long attention span |
| 6. needs long explanation | 1 | 2 | 3 | 4 | 5 | 6 | 7 | catches on quickly |
| 7. often worried | 1 | 2 | 3 | 4 | 5 | 6 | 7 | seldom worried |
| 8. stubborn | 1 | 2 | 3 | 4 | 5 | 6 | 7 | cooperative |
| 9. unhappy | 1 | 2 | 3 | 4 | 5 | 6 | 7 | happy |
| 10. easily distracted | 1 | 2 | 3 | 4 | 5 | 6 | 7 | concentrates |
| 11. self-conscious | 1 | 2 | 3 | 4 | 5 | 6 | 7 | comfortable with people |
| 12. doesn't know much | 1 | 2 | 3 | 4 | 5 | 6 | 7 | knows a lot |
| 13. high strung | 1 | 2 | 3 | 4 | 5 | 6 | 7 | easy going |
| 14. quarrelsome | 1 | 2 | 3 | 4 | 5 | 6 | 7 | agreeable |
| 15. moody | 1 | 2 | 3 | 4 | 5 | 6 | 7 | even-tempered |
| 16. gives up quickly | 1 | 2 | 3 | 4 | 5 | 6 | 7 | sticks to it |
| 17. fearful | 1 | 2 | 3 | 4 | 5 | 6 | 7 | self-confident |
| 18. not too bright | 1 | 2 | 3 | 4 | 5 | 6 | 7 | smart |
| 19. anxious | 1 | 2 | 3 | 4 | 5 | 6 | 7 | calm |
| 20. easily angered | 1 | 2 | 3 | 4 | 5 | 6 | 7 | good-natured |
| 21. complaining | 1 | 2 | 3 | 4 | 5 | 6 | 7 | contented |
| 22. careless about work | 1 | 2 | 3 | 4 | 5 | 6 | 7 | does careful work |

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for \$4.00 each, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be pre-paid. Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

The publications listed here are only those that have been prepared since 1982. For a complete, annotated list of all IRLD publications, write to the Editor.

- Wesson, C., Mirkin, P., & Deno, S. Teachers' use of self instructional materials for learning procedures for developing and monitoring progress on IEP goals (Research Report No. 63). January, 1982.
- Fuchs, L., Wesson, C., Tindal, G., Mirkin, P., & Deno, S. Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). January, 1982.
- Potter, M., & Mirkin, P. Instructional planning and implementation practices of elementary and secondary resource room teachers: Is there a difference? (Research Report No. 65). January, 1982.
- Thurlow, M. L., & Ysseldyke, J. E. Teachers' beliefs about LD students (Research Report No. 66). January, 1982.
- Graden, J., Thurlow, M. L., & Ysseldyke, J. E. Academic engaged time and its relationship to learning: A review of the literature (Monograph No. 17). January, 1982.
- King, R., Wesson, C., & Deno, S. Direct and frequent measurement of student performance: Does it take too much time? (Research Report No. 67). February, 1982.
- Greener, J. W., & Thurlow, M. L. Teacher opinions about professional education training programs (Research Report No. 68). March, 1982.
- Algozzine, B., & Ysseldyke, J. Learning disabilities as a subset of school failure: The oversophistication of a concept (Research Report No. 69). March, 1982.
- Fuchs, D., Zern, D. S., & Fuchs, L. S. A microanalysis of participant behavior in familiar and unfamiliar test conditions (Research Report No. 70). March, 1982.

- Shinn, M. R., Ysseldyke, J., Deno, S., & Tindal, G. A comparison of psychometric and functional differences between students labeled learning disabled and low achieving (Research Report No. 71). March, 1982.
- Thurlow, M. L., Graden, J., Greener, J. W., & Ysseldyke, J. E. Academic responding time for LD and non-LD students (Research Report No. 72). April, 1982.
- Graden, J., Thurlow, M., & Ysseldyke, J. Instructional ecology and academic responding time for students at three levels of teacher-perceived behavioral competence (Research Report No. 73). April, 1982.
- Algozzine, B., Ysseldyke, J., & Christenson, S. The influence of teachers' tolerances for specific kinds of behaviors on their ratings of a third grade student (Research Report No. 74). April, 1982.
- Wesson, C., Deno, S., & Mirkin, P. Research on developing and monitoring progress on IEP goals: Current findings and implications for practice (Monograph No. 18). April, 1982.
- Mirkin, P., Marston, D., & Deno, S. L. Direct and repeated measurement of academic skills: An alternative to traditional screening, referral, and identification of learning disabled students (Research Report No. 75). May, 1982.
- Algozzine, B., Ysseldyke, J., Christenson, S., & Thurlow, M. Teachers' intervention choices for children exhibiting different behaviors in school (Research Report No. 76). June, 1982.
- Tucker, J., Stevens, L. J., & Ysseldyke, J. E. Learning disabilities: The experts speak out (Research Report No. 77). June, 1982.
- Thurlow, M. L., Ysseldyke, J. E., Graden, J., Greener, J. W., & Mecklenberg, C. Academic responding time for LD students receiving different levels of special education services (Research Report No. 78). June, 1982.
- Graden, J. L., Thurlow, M. L., Ysseldyke, J. E., & Algozzine, B. Instructional ecology and academic responding time for students in different reading groups (Research Report No. 79). July, 1982.
- Mirkin, P. K., & Potter, M. L. A survey of program planning and implementation practices of LD teachers (Research Report No. 80). July, 1982.
- Fuchs, L. S., Fuchs, D., & Warren, L. M. Special education practice in evaluating student progress toward goals (Research Report No. 81). July, 1982.
- Kuehnle, K., Deno, S. L., & Mirkin, P. K. Behavioral measurement of social adjustment: What behaviors? What setting? (Research Report No. 82). July, 1982.

- Fuchs, D., Dailey, Ann Madsen, & Fuchs, L. S. Examiner familiarity and the relation between qualitative and quantitative indices of expressive language (Research Report No. 83). July, 1982.
- Videen, J., Deno, S., & Marston, D. Correct word sequences: A valid indicator of proficiency in written expression (Research Report No. 84). July, 1982.
- Potter, M. L. Application of a decision theory model to eligibility and classification decisions in special education (Research Report No. 85). July, 1982.
- Greener, J. E., Thurlow, M. L., Graden, J. L., & Ysseldyke, J. E. The educational environment and students' responding times as a function of students' teacher-perceived academic competence (Research Report No. 86). August, 1982.
- Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study (Research Report No. 87). August, 1982.
- Skiba, R., Wesson, C., & Deno, S. L. The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 88). September, 1982.
- Marston, D., Tindal, G., & Deno, S. L. Eligibility for learning disability services: A direct and repeated measurement approach (Research Report No. 89). September, 1982.
- Thurlow, M. L., Ysseldyke, J. E., & Graden, J. L. LD students' active academic responding in regular and resource classrooms (Research Report No. 90). September, 1982.
- Ysseldyke, J. E., Christenson, S., Pianta, R., Thurlow, M. L., & Algozzine, B. An analysis of current practice in referring students for psycho-educational evaluation: Implications for change (Research Report No. 91). October, 1982.
- Ysseldyke, J. E., Algozzine, B., & Epps, S. A logical and empirical analysis of current practices in classifying students as handicapped (Research Report No. 92). October, 1982.
- Tindal, G., Marston, D., Deno, S. L., & Germann, G. Curriculum differences in direct repeated measures of reading (Research Report No. 93). October, 1982.
- Fuchs, L.S., Deno, S. L., & Marston, D. Use of aggregation to improve the reliability of simple direct measures of academic performance (Research Report No. 94). October, 1982.
- Ysseldyke, J. E., Thurlow, M. L., Mecklenburg, C., & Graden, J. Observed changes in instruction and student responding as a function of referral and special education placement (Research Report No. 95). October, 1982.

- Fuchs, L. S., Deno, S. L., & Mirkin, P. K. Effects of frequent curriculum-based measurement and evaluation on student achievement and knowledge of performance: An experimental study (Research Report No. 96). November, 1982.
- Fuchs, L. S., Deno, S. L., & Mirkin, P. K. Direct and frequent measurement and evaluation: Effects on instruction and estimates of student progress (Research Report No. 97). November, 1982.
- Tindal, G., Wesson, C., Germann, G., Deno, S. L., & Mirkin, P. K. The Pine County model for special education delivery: A data-based system (Monograph No. 19). November, 1982.
- Epps, S., Ysseldyke, J. E., & Algozzine, B. An analysis of the conceptual framework underlying definitions of learning disabilities (Research Report No. 98). November, 1982.
- Epps, S., Ysseldyke, J. E., & Algozzine, B. Public-policy implications of different definitions of learning disabilities (Research Report No. 99). November, 1982.
- Ysseldyke, J. E., Thurlow, M. L., Graden, J. L., Wesson, C., Deno, S. L., & Algozzine, B. Generalizations from five years of research on assessment and decision making (Research Report No. 100). November, 1982.
- Marston, D., & Deno, S. L. Measuring academic progress of students with learning difficulties: A comparison of the semi-logarithmic chart and equal interval graph paper (Research Report No. 101). November, 1982.
- Beattie, S., Grise, P., & Algozzine, B. Effects of test modifications on minimum competency test performance of third grade learning disabled students (Research Report No. 102). December, 1982.
- Algozzine, B., Ysseldyke, J. E., & Christenson, S. An analysis of the incidence of special class placement: The masses are burgeoning (Research Report No. 103). December, 1982.
- Marston, D., Tindal, G., & Deno, S. L. Predictive efficiency of direct, repeated measurement: An analysis of cost and accuracy in classification (Research Report No. 104). December, 1982.
- Wesson, C., Deno, S., Mirkin, P., Sevcik, B., Skiba, R., King, R., Tindal, G., & Maruyama, G. Teaching structure and student achievement effects of curriculum-based measurement: A causal (structural) analysis (Research Report No. 105). December, 1982.
- Mirkin, P. K., Fuchs, L. S., & Deno, S. L. (Eds.). Considerations for designing a continuous evaluation system: An integrative review (Monograph No. 20). December, 1982.
- Marston, D., & Deno, S. L. Implementation of direct and repeated measurement in the school setting (Research Report No. 106). December, 1982.

- Deno, S. L., King, R., Skiba, R., Sevcik, B., & Wesson, C. The structure of instruction rating scale (SIRS): Development and technical characteristics (Research Report No. 107). January, 1983.
- Thurlow, M. L., Ysseldyke, J. E., & Casey, A. Criteria for identifying LD students: Definitional problems exemplified (Research Report No. 108). January, 1983.
- Tindal, G., Marston, D., & Deno, S. L. The reliability of direct and repeated measurement (Research Report No. 109). February, 1983.
- Fuchs, D., Fuchs, L. S., Dailey, A. M., & Power, M. H. Effects of pre-test contact with experienced and inexperienced examiners on handicapped children's performance (Research Report No. 110). February, 1983.
- King, R. P., Deno, S., Mirkin, P., & Wesson, C. The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 111). February, 1983.
- Tindal, G., Deno, S. L., & Ysseldyke, J. E. Visual analysis of time series data: Factors of influence and level of reliability (Research Report No. 112). March, 1983.
- Tindal, G., Shinn, M., Fuchs, L., Fuchs, D., Deno, S., & Germann, G. The technical adequacy of a basal reading series mastery test (Research Report No. 113). April, 1983.
- Sevcik, B., Skiba, R., Tindal, G., King, R., Wesson, C., Mirkin, P., & Deno, S. Communication of IEP goals and student progress among parents, regular classroom teachers, and administrators using systematic formative evaluation (Research Report No. 114). April, 1983.
- Wesson, C. Two student self-management techniques applied to data-based program modification (Research Report No. 115). April, 1983.
- Wesson, C., Skiba, R., Sevcik, B., King, R., Tindal, G., Mirkin, P., & Deno, S. The impact of the structure of instruction and the use of technically adequate instructional data on reading improvement (Research Report No. 116). May, 1983.
- Wesson, C. Teacher vs student selection of instructional activities (Research Report No. 117). May, 1983.
- Tindal, G., & Deno, S. Factors influencing the agreement between visual and statistical analyses of time series data (Research Report No. 118). June, 1983.
- Skiba, R. S. Classroom behavior management: A review of the literature (Monograph No. 21), June, 1983.
- Graden, J. L., Thurlow, M. L., & Ysseldyke, J. E. When are students most academically engaged? Academic responding time in different instructional ecologies (Research Report No. 119). June, 1983.

- Fuchs, L. S., Deno, S. L., & Roettger, A. The effect of alternative data-utilization rules on spelling achievement: An n of 1 study (Research Report No. 120). June, 1983.
- Skiba, R., Sevcik, B., Wesson, C., King, R., & Deno, S. The non-effect of process-product variables in resource classrooms (Research Report No. 121). June, 1983.
- Fuchs, L. Tindal, G., Shinn, M., Fuchs, D., Deno, S., & Germann, G. Technical adequacy of basal readers' mastery tests: The Ginn 720 series (Research Report No. 122). June, 1983.
- Tindal, G., Germann, G., Marston, D., & Deno, S. The effectiveness of special education: A direct measurement approach (Research Report No. 123). June, 1983.
- Sevcik, B., Skiba, R., Tindal, G., King, R., Wesson, C., Mirkin, P., & Deno, S. Curriculum-based measurement: Effects on instruction, teacher estimates of student progress, and student knowledge of performance (Research Report No. 124). July, 1983.
- Skiba, R., Marston, D., Wesson, C., Sevcik, B., & Deno, S. L. Characteristics of the time-series data collected through curriculum-based reading measurement (Research Report No. 125). July, 1983.
- Ysseldyke, J., Christenson, S., Graden, J., & Hill, D. Practical implications of research on referral and opportunity to learn (Monograph No. 22). July, 1983.
- Marston, D., Deno, S., & Tindal, G. A comparison of standardized achievement tests and direct measurement techniques in measuring pupil progress (Research Report No. 126). July, 1983.
- Fuchs, D., Fuchs, L. S., Tindal, G., & Deno, S. L. Variability of performance: A "signature" characteristic of learning disabled children? (Research Report No. 127). July, 1983.
- Tindal, G., Fuchs, L., Fuchs, D., Shinn, M., Deno, S., & Germann, G. The technical adequacy of a basal series mastery test: The Scott-Foresman reading program (Research Report No. 128). July, 1983.
- Fuchs, L. S., Fuchs, D., & Deno, S. L. The nature of inaccuracy among readability formulas (Research Report No. 129). July, 1983.
- Fuchs, L., Tindal, G., Fuchs, D., Shinn, M., Deno, S., & Germann, G. The technical adequacy of a basal reading mastery test: The Holt basic reading series (Research Report No. 130). July, 1983.
- Ysseldyke, J. E., Christenson, S., Algozzine, B., & Thurlow, M. L. Classroom teachers' attributions for students exhibiting different behaviors (Research Report No. 131). July, 1983.

Tindal, G., Germann, G., & Deno, S. Descriptive research on the Pine County norms: A compilation of findings (Research Report No. 132). July, 1983.

Skiba, R. J. The relationship between classroom management strategies and student misbehaviors (Research Report No. 133). July, 1983.

Fuchs, D., Fuchs, L. S., Power, M. H., & Dailey, A. M. Systematic bias in the assessment of handicapped children (Research Report No. 134). July, 1983.

Fuchs, D., & Fuchs, L. S. The importance of scorer bias to handicapped preschoolers' stronger performance with familiar examiners (Research Report No. 135). July, 1983.